Title Page

Manuscript title: Advancing the Next Generation of Risk Assessment

Ila Cote,[1] Melvin Andersen,[2] Gerald Ankley,[1] Stanley Barone, [1] Linda Birnbaum,[3] Kim Boekelheide,[4] Frederic Bois,[5] Lyle Burgoon,[1] Weihsueh Chiu,[6] Douglas Crawford-Brown,[7] Kevin Crofton,[1] Michael DeVito,[3] Robert Devlin,[1] David Dix,[1] Stephen Edwards,[1] Kathrine Guyton,[8] Dale Hattis,[9] Richard Judson,[1] Derek Knight,[10] Daniel Krewski,[11] Jason Lambert,[1] Elizabeth Maull,[3] Donna Mendrick,[12] Stephanie Padilla,[1] Greg Paoli,[13] Chirag Patel,[14] Edward Perkins,[15] Gerald Poje,[16] Christopher Portier,[17] Ivan Rusyn,[6] Paul Schulte,[18] Anton Simeonov,[19] Martyn Smith,[20] Christine Sofge,[18] Kristine Thayer,[3] Russell Thomas,[1] Reuben Thomas,[20] Ray Tice,[3] John Vandenberg,[1] Dan Villanueve,[1] Scott Wesselkamper,[1] Maurice Whelan,[21] Ron White[22] Menghang Xia,[19] Carole Yauk,[23] Lauren Zeise,[24] Jay Zhao,[1] Robert DeWoskin[1]

[1]U.S. Environmental Protection Agency
[2]The Hamner Institutes for Health Sciences
[3]National Institute of Environmental Health Sciences and the National Toxicology Program
[4]Brown University
[5]L'Institut National de l'Environnement Industriel et des Risques
[6]Texas A & M University
[7]University of Cambridge, Department of Land Economy,
[8]International Agency for Cancer Research
[9]Clark University
[10]European Chemicals Agency
[11]University of Ottawa
[12]Food and Drug Administration, National Center for Toxicological Research
[13]Risk Sciences International
[14] Harvard University
[15]Army Corps of Engineers
[16]Grant Consulting Group
[17]Environmental Defense Fund
[18]National Institute for Occupational Safety and Health
[19]National Institutes of Health, National Center for Advancing Translational Science
[20]University of California, Berkeley
[21]European Joint Research Commission
[22]Center for Effective Government
[23]Health Canada
[24]California Environmental Protection Agency, Office of Environmental Health Hazard Assessment

Corresponding author: Ila Cote, U.S. Environmental Protection Agency, Region 8, Room 8152, 1595 Wynkoop St., Denver, Colorado, USA 80202-1129. Telephone: 202-288-9539. Email: cote.ila@epa.gov.

Running title: Advancing the Next Generation of Risk Assessment

Acknowledgments, including grant information

Disclaimer: The views in this article are those of the authors, and do not necessarily reflect the views or policies of U.S. EPA.

The authors declare they have no actual or potential competing financial interests.

Abstract

Background and Objectives

The Next Generation (NexGen) of Risk Assessment effort is a multiyear, multimillion-dollar collaboration by U.S. federal and state regulatory agencies, international agencies, academic institutions, and the nonprofit sector to evaluate new, potentially more efficient approaches to environmental health risk assessment. NexGen was motivated by recent advances in molecular, computational, and systems biology that might augment or replace existing risk assessment methods. The purpose of this paper is to provide the key findings of the NexGen program and to identify strategic directions for additional research.

Methods

Central to the NexGen effort was to evaluate how new data, methods and models might support different types of decisions that risk managers must make. We developed prototypes demonstrating application of new data and methods to decisions that had increasing levels of regulatory impact. Data types included transcriptomics, genomics, proteomics, nuclear and cell receptor assays, and various cell toxicity markers. Methods included molecular epidemiology and clinical studies, bioinformatic knowledge mining, short-duration in vivo bioassays, quantitative structure activity relationship approaches, and high-throughput in vitro bioassays.

Conclusions

NexGen has fostered extensive discussion in the risk science and management communities and advanced our ability to apply new science to better assess potential public health risks. EPA in collaboration with other federal agencies, are developing more specific prototype risk assessments informed by advanced biology that are anticipated to be more efficient that existing methods.

Introduction

Background

Advances in molecular and cell biology provide new insights into the causes and risk factors associated with disease, largely by evaluating molecular events that influence cell functions and interactions. High-throughput/high-content (HT/HC) assays and robotic implementation are generating large data streams at unprecedented speeds. Computational tools, automated analytical methods (bioinformatics), and systems biology approaches are being developed to organize and interpret the information. Risk assessment and toxicity testing are poised to benefit greatly from these advances.

The National Library of Medicine (NLM) and others are compiling, organizing, managing, and storing these data in support of efforts to understand public health determinants better and to help answer such questions as: Which chemicals are environmentally better choices? Why do individuals and specific subpopulations respond differently to chemical exposures? What happens when people are exposed to low levels of chemicals and to multiple chemicals? How do other factors like poverty and preexisting illness influence public health risk? How might evaluation and application of these data, methods and models be used to support environmental health decision making?

To evaluate how new data types and approaches might enhance environmental

health risk assessments, the U.S. Environmental Protection Agency (EPA) collaborated with many U.S. and international organizations (Table 1) to consider the state of science and to develop case studies (illustrative prototypes) demonstrating various approaches that investigators could apply, singly or in combination, to different risk management problems. Our goal was not to evaluate all data and all risk management situations but to provide examples that could promote discussion in the risk assessment, risk management, and stakeholder communities, and that would facilitate the transition from strategy to practical application. This paper summarizes these efforts.

Objectives

Our specific objectives were to test whether we can identify specific patterns of molecular events that (1) are strongly associated with the adverse effects of chemical exposures; (2) are exposure-dose dependent within the range of environmental exposures; (3) can be shown to vary with risk factors such as genomic variants, mixture, and nonchemical stressor exposures; and (4) can be used as indicators of adverse health effects and chemical potency. Additionally, we wanted to consider how new types of assessments might address differing risk management needs or risk context and to develop decision rules for integrating and applying the available data.

Methods

We evaluated and integrated diverse types of data to evaluate if, and how, we can use advanced biological data to better inform risk assessment.

Preparation for Prototype Development

We established the foundation for this effort by: (1) identifying EPA risk managers' needs and developing a strategy for the overall approach (Cote et al. 2012); (2) consulting with experts on the concepts for the prototypes (EPA 2010); (3) holding a conference to inform the public about upcoming activities and to solicit advice (EPA 2011); and (4) developing a framework that articulated the guiding principles for the NexGen effort (Krewski et al. 2014).

Risk Assessments Targeted to Various Decision Contexts

We developed seven prototypes that illustrate three decision contexts, which generally represent the types of environmental challenges risk managers face. The three decision contexts we defined are: (1) major scope, usually regulatory decision-making, generally aimed at nationwide exposures and associated risks; (2) limited scope, usually nonregulatory decision-making, generally aimed at more limited exposure, hazard, or data situations; and (3) screening and prioritization of chemicals for further testing, research, or assessment, or for emergency response (Figure 1). From left to right in Figure 1, the amount of traditional data available for assessment (e.g., in vivo rodent toxicity data, epidemiology data) and the confidence in the assessment conclusions decreases but the number of chemicals that can be evaluated increases. As a caveat, these decision contexts are generalized and do not capture all decisions risk managers face nor do they address the nuances of each situation.

Systematic Review and Criteria for Study Selection

Systematic review and criteria for study selection help ensure reproducibility,

transparency, and scientific acceptability of the NexGen assessment prototypes (DHHS 2014; Meek et al. 2014; NRC 2014; Rhomberg et al. 2013). Study selection criteria are similar to those used for traditional data but we augmented them with criteria specifically applicable to new methodologies (e.g., Minimum Information About a Microarray Experiment standard) (McConnell et al. 2014). Rapidly evolving best practices for advanced biology and certain reporting requirements led many initially considered studies to be deemed inadequate for risk assessment purposes.

The Prototypes

We considered a variety of methods in the prototypes, which are summarized in Table 2 (EPA 2014b; Krewski et al. 2014). The following describes the results from the prototypes developed in the three illustrative decision-context categories.

Major-scope Assessment Prototypes

Major-scope prototypes explored how toxicogenomic studies of exposed human populations can inform risk assessment by characterizing early events in the cascade of events leading to adverse outcomes, biomarkers of exposure and effects, factors contributing to population variability and susceptibility, and the low exposure-response relationship (McHale et al. 2011). We developed these prototypes primarily to explore proof of concept and secondarily as examples of how new data types could inform (or be consistent with) chemical assessments already based on a robust traditional data set.

We evaluated transcriptomic molecular data (epidemiological or clinical) in the range of environmental exposures for three environmental chemicals: (1) benzene and

other leukemogens (McHale et al. 2011; McHale et al. 2012; Smith et al. 2011; R Thomas et al. 2012; R Thomas et al. 2013; R Thomas et al. 2014); (2) ozone (EPA 2013a; Hatch et al. 2014; McCullough et al. 2014); and (3) polycyclic aromatic hydrocarbons (PAHs), including tobacco smoke and benzo[a]pyrene (DHHS 2014; EPA 2013b)). We also considered genomic, proteomic, and epigenomic data as available. We evaluated exposures for benzene of <0.1 to 10 parts per million (ppm) and ozone of 0.5 ppm for 2 hours. We used precise individual measures of exposure-dose for benzene and ozone (benzene urinary metabolites and $^{18}O_2$) (Hatch et al. 2014; Vermeulen et al. 2004). For PAHs exposures we used self-reported smoking We also considered animal molecular data for B[a]P (EPA 2013b). The PAH/tobacco smoke prototype differed from the benzene and ozone efforts by focusing on pathway mining of existing human microarray data from the Gene Expression Omnibus and ArrayExpress (NCBI 2014, EMBL-EBI, 2014). This prototype demonstrated how the data could be used to evaluate whether smokers have a larger number of gene expression changes than nonsmokers that are similar to the gene expression changes associated with lung cancer.

The prototypes focused on toxicogenomics anchored qualitatively and quantitatively to known health outcomes associated with these chemicals, that is, hematotoxicity and leukemia (benzene), lung inflammation and injury (ozone), and lung cancer (PAHs). These data-rich associations enabled us to draw on a wealth of chemical- and disease-specific data to interpret the molecular biology findings.

Limited-scope Assessment Prototypes

Limited-scope assessment prototypes explored approaches that fall between molecular human clinical and epidemiology studies (above) and in vitro, HT screening bioassays (below) in terms of confidence in the data to characterize public health risks, resources expended to collect data, and the number of chemicals that can be evaluated in a given period. We considered three approaches: (1) knowledge mining of large health databases (focusing on human tissue biomonitoring and diabetes data from NHANES [National Health and Nutrition Examination Survey] data) (Bell and Edwards 2014; EPA 2014b; Patel et al. 2012; Patel et al. 2013a; Thayer et al. 2012); (2) short-duration, *in vivo* exposures using alternative (nonmammalian) species (largely focusing on the thyroid hormone disruptor mechanism, and zebrafish developmental outcomes for several hundred chemicals) (Padilla et al. 2012; Perkins et al. 2013; Sipes et al. 2011a; Sipes et al. 2011b; Thienpont et al. 2011; Villeneuve et al. 2014); and (3) short-duration, *in vivo* exposure rodent studies that correlated transcriptomic alterations following exposure to 11 chemicals with cancer and noncancer outcomes as determined in traditional bioassays (RS Thomas et al. 2011; RS Thomas et al. 2012; RS Thomas et al. 2013a, RS Thomas 2013c). Advantages of the limited-scope approaches compared to HT *in vitro* approaches include intact metabolism, intact cell and tissue interactions, and the potential to measure adverse health outcomes, including complex outcomes such as altered behavior and development.

Screening and Prioritization

The two screening and prioritization prototypes are (1) quantitative structure activity relationship (QSAR) models and use of analogous chemicals to expand available

information (also called "read-across") (EPA 2014c; Golbraikh et al. 2012; OECD 2014; Politi et al. 2014; Wang et al. 2011; Wang et al. 2012a); and (2) *in vitro*, cell-free, enzymatic and ligand-binding HT screening assays and modeling approaches as illustrated by EPA's National Center for Computational Toxicology program (DeWoskin et al. 2014; Judson 2010; Judson et al. 2011; Judson et al. 2012; Judson et al. 2013; Judson et al. 2014; Kavlock et al. 2012; Kleinstreuer et al. 2014; Knudsen and DeWoskin 2011; Knudsen et al. 2013; Sipes et al. 2013) and the multiagency collaborative Tox21 program (Attene-Ramos et al. 2013; Tice et al. 2013). The second prototype focuses on evaluating thyroid hormone disruptors (Cox et al. 2014; Rotroff et al. 2013; Sipes et al. 2011a).

Examining Human Variability in Responses

Evaluating human variability in response to environmental factors and protecting vulnerable population segments is an often-stated goal. Unfortunately, the data to evaluate variability and susceptibility are generally scant. For the NexGen effort, we evaluated several data types to inform this issue: (1) adverse outcome networks to identify mechanistic commonalties among leukemogens and lifestyle factors (diet and stress) that alter leukemia risks (EPA 2014b; IARC 2012; Smith et al. 2011; R Thomas et al. 2012); (2) altered disease incidence in population segments that have specific genetic polymorphisms (EPA 2014b); (3) data for *in vitro* cells that retain an asthma phenotype in ozone studies (Duncan et al. 2012); (4) correlated measures of phenotypic differences among diverse subpopulations with expression patterns (EPA 2014b; Patel et al. 2012; Patel et al. 2013a); (5) HT *in vitro* data from cell lines with different genetic backgrounds

from the 1000 Genomes effort (Lock et al. 2012; O'Shea et al. 2011); and (6) computational modeling in which variability in parameter values is simulated for differences among subpopulations (Knudsen and DeWoskin 2011; Shah and Wambaugh 2010). See Zeise et al. (2013) for more details.

Results and Discussion

The NexGen prototypes demonstrate remarkable progress in our understanding of health and disease, and realization of the National Research Council's (NRC) vision embodied in *Toxicity Testing in the 21st Century* (NRC 2007). In the few years since NRC published that report, toxicity testing and risk assessment have begun shifting from the traditional use of animal data to using the new approaches the prototypes demonstrate. The new approaches consider a distinct and broader data array, foster mechanistic understanding of adversity, and move toward replacing uncertainty factors and extrapolations with data-derived probability distributions.

In each decision context category, new methods and data types were identified that could help inform assessment efforts. Methods illustrated in the screening and prioritization (Tier 1) and limited-scope (Tier 2) prototypes originally were designed for qualitative evaluation of chemicals. Already, however, new and integrated approaches are being developed to estimate relative potencies and more rapid quantitative toxicity values for use in certain decision contexts.

AOPs were used extensively to organize and interpret data for most of the prototypes and are viewed as extremely important in terms of linking molecular events to apical outcomes. The concept of AOPs and networks has gained considerable traction since it was first introduced (Ankley et al. 2010; Kleensang et al. 2014; NAS 2012; Tollefsen et

al. 2014; Vinken 2013). (Mechanism of action, mode of action, toxicity pathway and adverse outcome pathway are all terms used to describe causal events leading to toxicity and disease. The term AOP is not ideal as it sometimes erroneously conveys that toxicity results from novel events rather than perturbations of normal biology but AOP or AOP network is used throughput this paper due to its common use among a number of US and European Agencies.)

Data quality and reporting are significant issues going forward. Our data searches to develop the prototypes identified many published studies that we could not use because either the data or the reporting did not adequately meet the criteria for use in health risk assessment. This situation results in part from the lag before best practices are developed and fully implemented in the research community and inconsistent application of criteria for data quality and reporting (EPA 2014b, McConnell et al 2014).

Integrating the available data into a coherent analysis is also a challenge. Rhomberg et al. (2013) reviewed 50 existing "weight-of-evidence" frameworks (termed evidence integration by NRC, 2014). They identified four phases of analysis consistently used in the 50 frameworks: "(1) defining the causal question and developing criteria for study selection, (2) developing and applying criteria for review of individual studies, (3) evaluating and integrating evidence and (4) drawing conclusions based on inferences" (Rhomberg et al. 2013). Steps 1 and 2, as used in the U.S. federal government, are discussed in some detail in DHHS (2014), McConnell et al. (2014), and NRC (2014). Supplemental Table 2 presents an "Illustrative Framework for Evidence Integration for New Data Types," focusing on Steps 3 and 4—evaluating and integrating evidence and

drawing conclusions based on inferences. Supplemental Table 2 draws on previous works for the basis of evidence integration (DHHS 2014; EPA 2013b; Meek et al. 2014; NRC 2014).

Major-scope Assessment Prototypes (Tier 3)

We designed the Tier 3 prototypes to test the hypothesis that new data types could provide qualitative and quantitative results comparable to those that robust traditional data. Secondarily, we evaluated if new data types could add to information provided by robust traditional data sets. We discuss support for this hypothesis and several sources of variability below (EPA 2013a, 2013b, 2014b; Esposito et al. 2014: Hatch et al. 2014; McCullough et al. 2014; McHale et al. 2011; McHale et al. 2012; Smith et al. 2011; Thomas et al. 2014; Van Dyck et al 2014; Wang X et al. 2014).

- AOP networks appear useful in predicting specific hazards for benzene and other known leukemogens (hematotoxicity), ozone (lung inflammation and injury), and PAHs (lung cancer), as well as providing quantitative biomarkers. Chemical and nonchemical stressors appear to perturb various pathways within the same disease associated network, but do not always affect the same expressed genes or pathway. Hence, overly simplistic descriptions of AOPs likely miss the potential for network-level interactions. The observed network modifications evaluated appear causally related to specific adverse effects. Evidence for causality included pharmacologic intervention to block identified pathway changes and concomitant amelioration of severity or incidence of specific adverse outcomes. We concluded that less well-studied chemicals inducing the same AOP or AOP network could be of concern for similar health outcomes. Conversely, lack

of an apparent mechanistic link to an adverse outcome could be a rationale for downgrading questionable *in vivo* data. Thus, we anticipate network level knowledge often will be highly valuable to understand causal mechanisms, integrate evidence, assess potential hazards less well-studied chemicals pose, and provide a method for cumulative assessment by grouping chemical and nonchemical stressors according to their common AOP network. As the prototypes illustrate, AOP networks also can help us evaluate the roles of human gene variants in subpopulation susceptibility or resistance (EPA 2014b).

- AOP component biomarkers, can help inform exposure-dose-response relationships, as shown in the benzene and ozone prototypes (and the Tier 2 thyroid hormone disruption prototype discussed below) illustrate. For benzene and ozone, the AOPs appear to evolve with increasing exposures. For example, with benzene, gene and pathway alterations indicative of impaired immune function are present at all exposure levels evaluated (from <0.1 ppm to 10 ppm) but, at higher concentrations, AOPs characteristic of more severe toxicity (apoptosis and cell death) begin to emerge. Thus, data collection over a range of environmental concentrations remains important in evaluating new data types. Also, limited time-course post exposure data were available for ozone; various AOPs involved in lung injury evolved post exposure, demonstrating the potential dynamic nature of underlying mechanisms (EPA 2013b; McCollough et al. 2014). One of the most promising applications of exposure/effect biomarkers is the ability to measure events of interest directly in environmentally exposed humans; such applications are revolutionizing epidemiology.
- Chemical exposures resulting in diseases appear to share AOP networks with diseases of

unknown origins (idiopathic or potentially naturally occurring disease). Chemically induced adverse effects evaluated here appear to add to naturally occurring backgrounds of disease, via shared mechanisms (EPA 2014b). As NRC (2009) and Crump et al. (1976) discuss, this finding has implications for an assumption of low-dose linearity for cancer and noncancer outcomes at the population level.

- Across the prototypes, we observed that a variety of factors could introduce experimental variability, including exposure concentrations, time post exposure and dosimetry, differences in techniques (microarray vs. RNAseq), experimental paradigms (*in vivo* vs. *in vitro*, primary cell culture vs. cell lines), cell and tissue type, individual genomic profile, coexposures, and lifestage. Without tight control of variability identifying causal events correctly can be difficult even knowing the adverse outcome. This highlights the need for careful experimentation and interpretation when potential outcomes are unknown (EPA2014b).

Limited-scope Assessment Prototypes (Tier 2)

We designed the Tier 2 prototypes to evaluate data from knowledge-mining, alternative species tests and from short-term in vivo studies for identifying potential hazards, refining mechanistic understanding, and characterizing the relative potencies of hundreds to thousands of chemicals in a more rapid fashion than with traditional methods. Confidence in these data generally ranks between Tier 3 and Tier 1 approaches. Highlights from the prototypes are briefly discussed next.

- The limited scope approaches are faster and less expensive than the molecular

epidemiology and molecular clinical studies noted above. Furthermore, unlike the quantitative structure activity relationship (QSAR) models and HT screening data (discussed below), the data from *in vivo* studies are from intact systems for metabolism, normal architecture (for various cell types), and normal tissue interactions; and can be used to study more complex system-level outcomes, such as developmental and neurobehavioral outcomes.

- In the data-mining exercises specific chemical exposures were associated with altered diabetes or prediabetes risks (e.g., chlorinated organics, heavy metals, selected nutrients). Exposures were determined via NHANES human tissue biomonitoring, and incidence was clinically defined within NHANES. Additional potential risk factors – multiple chemicals exposures and genetic and lifestyle susceptibility traits – were also identified (Bell and Edwards, 2014; EPA 2014b, Patel et al. 2012; Patel et al. 2013a, 2013b). In one example, 59 percent of people with high levels of cadmium, lead, and arsenic also had markers for diabetes. The data mining results are generally most suitable for hypothesis generation because the output only identifies associations among events in very large data sets. The availability of biomonitoring data and clinical diagnosis in the same individuals, however, increase the weight of evidence for these data. Also, others have provided traditional and computational data supporting a link between chemical exposure and diabetes (Audouze et al. 2013; Dimas et al. 2014; Inadera 2013, Thayer et al. 2012).
- Two Tier 2 prototypes demonstrated use of short-duration exposures in alternative species and mammalian species, respectively. The results were evaluated with novel

molecular and computational approaches to provide insights into potential environmental risks. These short-duration exposure studies, in composite, successfully identified levels of exposures associated with key molecular events, AOP networks alterations, and adverse effects; provided useful data on complex mechanistic behaviors, effects of mixtures, species-to-species similarities and differences (Ankley and Gray 2013); and illustrated how these data could be used to evaluate potential hazards and chemical potencies.

Screening and Prioritization Prototypes (Tier 1)

For the first time in the history of risk assessment, new approaches can evaluate tens of thousands of chemicals relatively rapidly. Tens of thousands of chemicals covered by the European Regulation on Registration, Evaluation, Authorisation and Restriction of Chemicals Legislation are being evaluated using QSAR, and ~10,000 chemicals are being screened in the US Tox21 program using innovative robotic technology and in vitro bioassays. Kavlock et al. (2012) note that "These tools can probe chemical-biological interactions at fundamental levels, focusing on the molecular and cellular pathways that are targets of chemical disruption." The two prototypes, QSAR models (Goldsmith et al. 2012; Venkatapathy and Wang 2013; 2012a; Wang et al. 2012b) and HT *in vitro* bioassays, were used to illustrate rapid screening and prioritization of chemicals (Judson et al. 2013; Kavlock et al. 2012; Rusyn et al. 2012; Sipes et al. 2013; Tice et al. 2013). Additional insights include:

- An essential element to evaluation and application of HT data is characterization of dose.

Researchers are developing methods using reverse dosimetry to extrapolate concentrations delivered to *in vitro* test systems to the comparable doses for *in vivo* exposure to rodents (or other test species) or to humans (*in vitro*-to-*in vivo* extrapolation [IVIVE]) (Hubal 2009; Rotroff et al. 2010; Wetmore et al. 2012; Wetmore et al. 2013). IVIVE extrapolation supports quantitative comparisons of *in vitro* toxicity results with *in vivo* bioassay results for estimating dose-response following human exposures.

- QSAR, *in vitro,* and *in silico* methods, are proving very useful for screening and ranking large numbers of chemicals for further evaluation and assessment, and urgent response situations where traditional data are lacking. Given the current state of the science, estimates of human disease risks based exclusively on QSAR and *in vitro* HT screening generally are too uncertain, however, *in silico* models are improving our understanding of these data. Insights into underlying mechanisms of toxicity, and the factors that might contribute to the variability in response to chemical exposure, are also progressing from these data streams and increasing their utility for understanding risks (Lock et al. 2012).

Caveats Pertaining to New Data Types in Risk Assessment

Generally, much of the new toxicogenomic data is associative in nature, that is, exposure and adverse outcomes can be associated with hundreds to thousands of gene changes, not all of which are likely to be causal (Mendrick 2011). Associative data, at best, generally "suggest" a causal relationship between exposure and adverse health outcomes. Criteria to move from "suggestive" to "likely" causal include meta-analyses of multiple, independent studies yielding similar results, experimental evidence of

alterations in putative AOP networks with consequent health outcomes (such as pharmacological interventions, gene knock-in/-out studies, alterations in risks due to human gene variants in key pathways), or combinations of traditional and NexGen data. The prototypes demonstrated how different types of evidence in each decision support category might be characterized with respect to causality and evidence integration (Supplemental Table 2) (EPA 2013a, NRC 2014). Additionally:

- The metabolism of many chemicals often is instrumental in toxicity. That most HT *in vitro* test systems are not metabolically competent should be taken into account. Although researchers are evaluating various approaches to add metabolic capability, satisfactory solutions are not yet available. Consequently, although positive results can be informative, negative results should not be interpreted, at this time, as a lack of toxicity.

- Cell type, tissue, individual, subpopulation, species, and test system can affect how specific alterations in molecular events manifest as adverse outcomes or disease, even when the molecular signature is the same. This phenomenon likely is due, at least in part, to dosimetry, epigenomic differences, and genomic plasticity, which assessments should consider, as feasible.

- Molecular profiles appear to be both dose and time dependent. Predicting adverse outcomes therefore can be challenging based only on "snapshots" of biological events. Some signatures do appear stable over time, however, and might serve as reliable indicators of chronic outcomes (RS Thomas et al. 2013c).

- Adverse outcome arguments in support of a regulatory assessment cannot be made solely with gene expression data, as messenger ribonucleic acid expression levels

cannot be used to infer protein activity directly. These data could, however, be suitable for ranking and screening. Gene expression data also can be used in an assessment to complement other mechanistic data.

- Our current ability to monitor multiple molecular processes (genomics, transcriptomics, proteomics, and epigenomics) in a single study is very limited, primarily due to expense. This lack of biological integration limits our understanding.

- Only a few chemicals, represented in the literature at this time, have advance biology data adequate to support regulatory risk assessments, due primarily to experimental design and reporting issues. This limitation emphasizes the need for systematic data review.

- A major issue in using molecular data in risk assessment is how best to interpret those data to predict observable adverse effects in humans. For example, how do changes in molecular events affect cells, changes in cells affect tissues and organs, and changes in organs affect the whole body? Researchers are collecting large amounts of HT/HC screening data on molecular-level effects, and the body of information on diseases and disease outcomes is substantial. Very sparse chemical-specific data are available, however, on intermediate levels of organization and on the sequence of cellular-level disruption of normal biology to effects at higher organizational levels. Even so, tremendous strides are being made in generating disease-specific information.

- Characterizing population variability in response give the many sources of inherent biological variability (e.g. genetic, epigenetic variants) among individuals is a major challenge. A second challenge is that each particular health outcome-chemical

exposure pair involves combinations of these sources, evaluation of which might be compounded further by extrinsic factors (e.g. diet, psychosocial stressors, other exogenous chemical exposures). A third challenge is that different decision contexts present distinct needs regarding the identification—and extent of characterization—of interindividual variability in the human population. New approaches to examining sources of variability in responses include: (1) computational modeling approaches in which variability in parameter values is simulated and differences among subpopulations are explored (Diaz Ochoa et al. 2012; Knudsen and DeWoskin 2011; Shah and Wambaugh 2010); (2) HT *in vitro* data generation using cell lines with different genetic backgrounds from the 1000 Genomes effort (Lock et al. 2012; O'Shea et al. 2011); (3) *in vivo* studies in genetically diverse strains of rodents to identify genetic determinants of susceptibility (Harrill et al. 2012; NIEHS 2014b); (4) comprehensive scanning of gene coding regions in panels of diverse individuals to examine the relationships among environmental exposures, interindividual sequence variation in human genes, and population disease risks (Mortensen and Euling 2013; NIEHS 2014a); (5) genome-wide association studies to uncover genomic loci that might contribute to human risk of disease (Abecasis et al. 2012; Bush and Moore 2012; NHGRI 2014; Wright et al. 2012); and (6) association studies that correlate measures of phenotypic differences among diverse populations with expression patterns for groupings of genes based on coexpression (Friend 2013; Patel et al. 2012; Patel et al. 2013a; Weiss et al. 2012). New understanding of the contribution of epigenomics to disease is advancing rapidly with evaluation of changes such as differential methylation of deoxyribonucleic acid (Hansen et al. 2011; Rakyan et al. 2011; Teschendorff and Widschwendter 2012).

Verifying toxicity testing schemes and computational models is essential for using these new data and approaches for risk-based decisions. Central to this effort is a framework and criteria for determining whether the new data types are adequate for various types of decisions. The level of certainty needed in the data varies with their intended use because inaccurate results have increasing consequence and costs as decisions progress from screening, to further testing, to what safe levels are, to what regulatory actions should be taken (Crawford-Brown 2013). Traditional "validation" schemes that evaluate conventional assay and testing structures do not adequately address the potential uses of these new data and methods and would require years to implement. Thus, as the technology for rapid, efficient, robust hazard testing advances, the verification process also must advance, to ensure confidence in their use. Clear and transparent articulation of these decision considerations will be essential to the acceptance of, and support for, assessment results, and in the overall evidence integration

Research Needs

Filling the gaps in our understanding of complex chemical and biological interactions at different levels of biological organization requires advanced research programs and models. Specific areas include:

- reliable, predictive molecular indicators for a wide variety of chemicals and diseases to assess hazard and characterize exposure and dose-response;
- identification of the networked interactions among genes, proteins, cells, tissues, organs, individuals, and populations, and the sequence of events at different levels that can lead

to disease (AOP networks) (Hartung and McBride 2011; Kleensang et al. 2014; Leist et al. 2014; Patel et al. 2013b);

- an integrated understanding of how genes are expressed and how the resulting proteins interact to maintain the body;
- methods to group chemical and nonchemical stressors based on common AOPs to enable cumulative risk assessment;
- data and methods to adjust for interspecies differences when assessing potential human toxicity based on nonhuman toxicity data;
- data and methods to characterize dose-response curves quantitatively for low-dose exposures;
- methods for non-aqueous exposures to chemicals present as gases or as airborne particles;
- methods such as reverse toxicokinetics models to extrapolate concentrations used in cellular and cell-free systems to in vivo doses; and
- methods to assess individual human variability due to genetic differences, preexisting disease and exposure, or adaptive and compensatory capabilities; and methods to incorporate this variability into population-level risk assessment.

Several large, federal, integrated research efforts are ongoing in the United States and Europe to improve toxicity testing and risk assessment. These programs are developing new data, advanced tools, and innovative technologies to evaluate chemical toxicity, integrate scientific information into state-of-the-science risk assessment, and optimize confidence in risk

management decisions.

Conclusions

EPA has registered more than 80,000 chemicals for use in the United States, and over a thousand more are introduced every year (EPA 2014a). The overarching challenge to risk assessors is to obtain and interpret data for assessing these chemicals quickly and efficiently to support decisions to protect public health and the environment. This challenge includes: (1) producing safer chemicals; (2) tracking the movement of chemicals and their byproducts through the environment; (3) identifying the sources of chemical exposures; (4) understanding the critical biological processes and toxicity pathways by which chemicals cause disease; and (5) evaluating the contribution of environmental chemical exposure to the overall disease burden for the general and susceptible populations (EPA 2012). The prototypes presented in this report demonstrate how new data can be used to help address several of these issues.

Based on the lessons learned in the NexGen program and elsewhere, several new types of high- and medium-throughput assessments are being advanced. Table 3 shows how characteristics of "fit-for-purpose" assessments could be tailored to support three illustrative decision-context categories. The table lists potential uses for NexGen assessments, data sources and types in different assessment categories, exposure paradigms used, incorporation of toxicokinetics, use of traditional data, hazard characterization, potency metrics, inferences drawn about the causal associations between exposures and adverse outcomes, and the numbers of chemicals that can be assessed, and the time to conduct any given assessment. Currently ongoing:

- Thousands of chemicals having no or very limited traditional data will be analyzed based on their similarities in physical-chemical structure to known toxicants to estimate their toxicity (QSAR modeling); and rapid, robotically conducted *in vitro* bioassay data will be used to identify a chemical's potency to alter important biological processes as indicators of apical toxicity (e.g., ToxCast and Tox21 programs).

- Thousands of chemicals will be evaluated using computer-based analyses of new and existing data, extracted from the published literature and then stored in extensive databases, to develop knowledge about the potential toxicity of chemicals and the causes of disease. Iteratively analyzing so much data from so many sources was previously impossible but is becoming the norm (e.g. Comparative Toxicogenomic Database, BioSystems).

- Hundreds of chemicals will be evaluated using a variety of new methods, including concerted, mechanistic approaches to understand the cumulative effects of multiple chemical and nonchemical stressors pose.

Near-term efforts will include more case study examples for public input and peer review, *and more opportunities to solicit stakeholder comments and participation. We are developing* a verification process for new methods and data types that focus on integrating the evidence into various decision contexts for use by risk assessors. The goal is to increase confidence in these new approaches for use in risk assessment. We will continue to identify and highlight significant scientific gaps from prototype development to be addressed in future research planning.

Logistical and methodological challenges in interpreting and using these new data and methods in risk assessment remain significant. Regardless, we anticipate the new approaches demonstrated in the prototypes soon will have a variety of applications for risk managers within EPA and the risk assessment community at large, including identifying safer chemicals and processes and reducing risk from exposures to hazardous chemicals in the environment. Major chemical assessments, for the present will continue to be driven by traditional data but augmented by new data types. The reader is encouraged to frequent the Internet sites of EPA and other research programs to learn about the latest developments and progress toward planned objectives in this rapidly evolving science.

Lastly, historically difficult risk assessment questions that new and emerging knowledge are likely to inform include: Why do individual and specific populations respond differently to environmental exposures? Are children at greater or lesser risk for certain exposures and effects, and if so, why? What happens when people are exposed to low levels of chemicals and mixture of chemical? How might other environmental factors like poverty and preexisting health conditions alter the response to chemical exposures?

References

Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, et al. 2012. An integrated map of genetic variation from 1,092 human genomes. Nature 491:56-65.

Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD, et al. 2010. Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. Environmental Toxicology and Chemistry 29:730-741.

Ankley GT and Gray LE. 2013. Cross-species conservation of endocrine pathways: A critical analysis of Tier 1 fish and rat screening assays with 12 model chemicals. Environ Toxicol Chem.

Attene-Ramos MS, Miller N, Huang R, Michael S, Itkin M, Kavlock RJ, et al. 2013. The tox21 robotic platform for the assessment of environmental chemicals - from vision to reality. Drug Discov Today 18:716-723.

Audouze K, Brunak S, Grandjean P. 2013. A computational approach to chemical etiologies of diabetes. Scientific reports 3:2712.

Bell S, Edwards S. Data mining NHANES to identify environmental chemical and disease associations. In: Proceedings of the Proceedings of the 2014 SIAM International Conferecnce on Data Mining, 2014.Bush WS, Moore JH. 2012. Chapter 11: Genome-wide association studies. Public Library of Science Computational Biology 8.

Cote I, Anastas PT, Birnbaum LS, Clark RM, Dix DJ, Edwards SW, et al. 2012. Advancing the next generation of health risk assessment. Environmental Health Perspectives 120:1499-1502.

Cox LA, Popken D, Marty MS, Rowlands JC, Patlewicz G, Goyak KO, et al. 2014. Developing scientific confidence in HTS-derived prediction models: Lessons learned from an endocrine case study. Regul Toxicol Pharmacol 69:443-450.

Crawford-Brown D. 2013. The role of advanced biological data in the rationality of risk-based regulatory decisions. Journal of Environmental Protection 4:238-249.

Crump KS, Hoel DG, Langley CH, Peto R. 1976. Fundamental carcinogenic processes and their implications for low dose risk assessment. Cancer Res 36:2973-2979.

DeWoskin RS, Knudsen TB, Shah I. 2014. Virtual models:Academic Press.

Dimas AS, Lagou V, Barker A, Knowles JW, Mägi R, Hivert MF, et al. Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. Diabetes. 2014 Jun;63(6):2158-71.

DHHS. 2014. The health consequences of smoking - 50 years of progress. Rockville, MD:Office of the Surgeon General.

Diaz Ochoa JG, Bucher J, Pery AR, Zaldivar Comenges JM, Niklas J, Mauch K. 2012. A multi-scale modeling framework for individualized, spatiotemporal prediction of drug effects and toxicological risk. Front Pharmacol 3:204.

Duncan KE, Dailey LA, Carson JL, Hernandez ML, Peden DB, Devlin RB. 2012. Cultured basal airway epithelial cells from asthmatics display baseline gene expression profiles that differ from normal healthy cells and exhibit differential responses to ambient air pollution particles American Journal of Respiratory and Critical Care Medicine 185:A4291.

EMBL-EBI. 2014. ArrayExpress, Available: http://www.ebi.ac.uk/ [accessed Aug 29, 2014]

EPA. 2010. Advancing the next generation (nexgen) of risk assessment: The prototypes workshop. Research Triangle Park, NC.

EPA. 2011. Advancing the next generation (nexgen) of risk assessment: Public dialogue conference report. Washington, DC.

EPA. 2012. Chemical safety for sustainability strategic research action plan 2012-2016. Office of Research and Development.

EPA. 2013a. Integrated science assessment of ozone and related photochemical oxidants (final report). Washington, DC.

EPA. 2013b. Iris toxicological review of benzo[a]pyrene (public comment draft). Available: http://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=66193 [accessed August 29, 2014.

EPA. 2014a. Next generation risk assessment: Incorporation of recent advances in molecular, computational, and systems biology (final report). EPA/600/R-14/004. Washington, DC:U.S. Environmental Protection Agency.

EPA. 2014b. Basic information about the toxic substance control act inventory. Available: http://www.epa.gov/oppt/existingchemicals/pubs/tscainventory/basic.html#background [accessed August 29, 2014.

EPA. 2014c. Quantitative structure activity relationship. Available: http://www.epa.gov/nrmrl/std/qsar/qsar.html [accessed August 29, 2014].

Esposito S. Air pollutants appear able to induce airway inflammation and increase... have demonstrated that exposure to common air pollutants plays a role in ....mechanisms linking air pollution and lung diseases in children, also taking ... BioMed Central 2014 www.biomedcentral.com/1471-2466/14/31.

Friend S. 2013. Scientific opportunities from heterogeneous biological data analysis: Overcoming complexity. Presentation. National Academy of Sciences Meeting: Integrating Environmental Health Data to Advance Discovery January 10-11.

Functional Genomics Data Society. 2014. Minimum Information About a Microarray Experiment. Available: http://fged.org/projects/miame/ [accessed August 29, 2014.

Golbraikh A, Wang XS, Zhu H, Tropsha A. 2012. Predictive qsar modeling: Methods and applications in drug discovery and chemical risk assessment. Handbook of Computational Chemistry:1309-1342.

Goldsmith MR, Peterson SD, Chang DT, Transue TR, Tornero-Velez R, Tan YM, et al. 2012. Informing mechanistic toxicology with computational molecular models. Methods in Molecular Biology 929:139-165.

Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG, et al. 2011. Increased methylation variation in epigenetic domains across cancer types. Nature Genetics 43:768-775.

Harrill AH, Desmet KD, Wolf KK, Bridges AS, Eaddy JS, Kurtz CL, et al. 2012. A mouse diversity panel approach reveals the potential for clinical kidney injury due to db289 not predicted by classical rodent models. Toxicological Sciences 130:416-426.

Hartung T, McBride M. 2011. Food for thought ... On mapping the human toxome. ALTEX 28:83-93.

Hatch GE, Duncan KE, Diaz-Sanchez D, Schmitt MT, Ghio AJ, Carraway MS, et al. 2014. Progress in assessing air pollutant risks from in vitro exposures: Matching ozone dose and effect in human airway cells. Toxicol Sci 141:198-205.

Hubal EA. 2009. Biologically relevant exposure science for 21st century toxicity testing. Toxicological Sciences 111:226-232.

IARC. 2012. Monograph 100f: Benzene.

Inadera H. 2013. Developmental origins of obesity and type 2 diabetes: Molecular aspects and role of chemicals. Environmental health and preventive medicine 18:185-197.

Judson R. 2010. In vitro screening of environmental chemicals for targeted testing prioritization: The toxcast project. Environmental Health Perspectives 118:485.

Judson R, Kavlock RJ, Setzer RW, Hubal EA, Martin MT, Knudsen TB, et al. 2011. Estimating toxicity-related biological pathway altering doses for high-throughput chemical risk assessment. Chemical Research in Toxicology 24:451-462.

Judson R, Martin MT, Egeghy P, Gangwal S, Reif DM, Kothiya P, et al. 2012. Aggregating data for computational toxicology applications: The U.S. Environmental Protection Agency (EPA) Aggregated Computational Toxicology Tesource (ACToR) system. International Journal of Molecular Science 13:1805-1831.

Judson R, Kavlock R, Martin M, Reif D, Houck K, Knudsen T, et al. 2013. Perspectives on validation of high-throughput assays supporting 21st century toxicity testing. Alternatives to Animal Experiments 30:51-56.

Judson R, Houck K, Martin M, Knudsen TB, Thomas RS, Sipes N, et al. 2014. In vitro and modelling approaches to risk assessment from the U>S. Environmental Protection Agency ToxCast programme. Basic & Clinical Pharmacology & Toxicology:n/a-n/a.

Kavlock R, Chandler K, Houck K, Hunter S, Judson R, Kleinstreuer N, et al. 2012. Update on EPA's ToxCast program: Providing high throughput decision support tools for chemical risk management. Chemical Research in Toxicology 25:1287-1302.

Kleensang A, Maertens A, Rosenberg M, Fitzpatrick S, Lamb J, Auerbach S, et al. 2014. T4 workshop report: Pathways of toxicity. Altex 31:53-61.

Kleinstreuer NC, Yang J, Berg EL, Knudsen TB, Richard AM, Martin MT, et al. 2014. Phenotypic screening of the toxcast chemical library to classify toxic and therapeutic mechanisms. Nature Biotechnology 32:583-591.

Knudsen T, DeWoskin RS. 2011. Systems modeling in development toxicity:John Wiley & Sons, Ltd.

Knudsen T, Martin M, Chandler K, Kleinstreuer N, Judson R, Sipes N. 2013. Predictive models and computational toxicology. Methods in Molecular Biology 947:343-374.

Krewski D, Westphal M, Andersen ME, Paoli GM, Chiu WA, Al-Zoughool M, et al. 2014. A framework for the next generation of risk science. Environ Health Perspect 122:796-805.

Leist M, Hasiwa N, Rovida C, Daneshian M, Basketter D, Kimber I, et al. 2014. Consensus report on the future of animal-free systemic toxicity testing. Altex 31:341-356.

Lock EF, Abdo N, Huang R, Xia M, Kosyk O, O'Shea SH, et al. 2012. Quantitative high-throughput screening for chemical toxicity in a population-based in vitro model. Toxicological Sciences 126:578-588.

McConnell ER, Bell SM, Cote I, Wang RL, Perkins EJ, Garcia-Reyero N, Gong P, Burgoon LD, 2014. Systematic Omics Analysis Review (SOAR) tool to support risk assessment. PLoS One 9:e110379. McCullough SD, Duncan KE, Swanton SM, Dailey LA, Diaz-Sanchez D, Devlin RB. 2014. Ozone induces a pro-inflammatory response in primary human bronchial epithelial cells through mitogen-activated protein kinase activation without nuclear factor-κb activation. Am J Respir Cell Mol Biol 51:426-435.

McHale CM, Zhang L, Lan Q, Vermeulen R, Li G, Hubbard AE, et al. 2011. Global gene expression profiling of a population exposed to a range of benzene levels. Environmental Health Perspectives 119:628-634.

McHale CM, Zhang L, Smith MT. 2012. Current understanding of the mechanism of benzene-induced leukemia in humans: Implications for risk assessment. Carcinogenesis 33:240-252.

Meek ME, Boobis A, Cote I, Dellarco V, Fotakis G, Munn S, et al. 2014. New developments in the evolution and application of the WHO/IPCS framework on mode of action/species concordance analysis. J Appl Toxicol 34:1-18.

Mendrick DL. 2011. Transcriptional profiling to identify biomarkers of disease and drug response. Pharmacogenomics 12:235-249.

Mortensen HM, Euling SY. 2013. Integrating mechanistic and polymorphism data to characterize human genetic susceptibility for environmental chemical risk assessment in the 21st century. Toxicology and Applied Pharmacology 271:395-404.

NAS. 2012. Emerging science for environmental health decisions; workshop − systems biology – informed risk assessment. National Academy of Sciences Building. 2101 Constitution Avenue, NW (between 21st and C Streets, NW). Washington, DC. June 14-15.

NCBI. 2014. Gene Expression Omnibus. Available: http://www.ncbi.nlm.nih.gov/geo/ [accessed August 29, 2014.

NHGRI. 2014. A catalog of published genome-wide association studies. Available: http://www.genome.gov/gwastudies/ [accessed August 29, 2014.

NIEHS. 2014a. NIEHS SNPs environmental genome project. Available: http://egp.gs.washington.edu/ [accessed August 29, 2014.

NIEHS. 2014b. Host susceptibility program. Available: http://grants.nih.gov/grants/guide/notice-files/NOT-ES-07-009.html [accessed August 29, 2014.

NRC. 2007. Toxicity testing in the 21st century: A vision and a strategy. Washington DC.

NRC. 2009. Science and decisions: Advancing risk assessment. Washington, DC.

NRC. 2014. Review of EPA's integrated risk information system (iris) process. Washington, D.C.:The National Academies Press.

O'Shea SH, Schwarz J, Kosyk O, Ross PK, Ha MJ, Wright FA, et al. 2011. In vitro screening for population variability in chemical toxicity. Toxicological Sciences 119:398-407.

OECD. 2014. The OECD QSAR toolbox. Version 3.0. Available: http://www.oecd.org/env/chemicalsafetyandbiosafety/assessmentofchemicals/theoecdqsartoolbox.htm [accessed August 29, 2014.

Padilla S, Corum D, Padnos B, Hunter DL, Beam A, Houck KA, et al. 2012. Zebrafish developmental screening of the toxcast phase i chemical library. Reproductive Toxicology 33:174-187.

Patel CJ, Chen R, Butte AJ. 2012. Data-driven integration of epidemiological and toxicological data to select candidate interacting genes and environmental factors in association with disease. Bioinformatics 28:i121-126.

Patel CJ, Chen R, Kodama K, Loannidis JP, Butte AJ. 2013a. Systematic identification of interaction effects between genome-and environment-wide associations in type 2 diabetes mellitus. Human Genetics 132(5):495-508.

Patel CJ, Sivadas A, Tabassum R, Preeprem T, Zhao J, Arafat D, et al. 2013b. Whole genome sequencing in support of wellness and health maintenance. Genome Medicine 5:58.

Perkins EJ, Ankley GT, Crofton KM, Garcia-Reyero N, Lalone CA, Johnson MS, et al. 2013. Current perspectives on the use of alternative species in human health and ecological risk assessments. Environmental Health Perspectives121(9):1002-10.

Politi R, Rusyn I, Tropsha A. 2014. Prediction of binding affinity and efficacy of thyroid hormone receptor ligands using qsar and structure-based modeling methods. Toxicol Appl Pharmacol 280:177-189.

Rakyan VK, Down TA, Balding DJ, Beck S. 2011. Epigenome-wide association studies for common human diseases. Nature Reviews Genetics 12:529-541.

Rhomberg LR, Goodman JE, Bailey LA, Prueitt RL, Beck NB, Bevan C, et al. 2013. A survey of frameworks for best practices in weight-of-evidence analyses. Critical Reviews in Toxicology 43:753-784.

Rotroff DM, Wetmore BA, Dix DJ, Ferguson SS, Clewell HJ, Houck KA, et al. 2010. Incorporating human dosimetry and exposure into high-throughput in vitro toxicity screening. Toxicological Sciences 117:348-358.

Rotroff DM, Dix DJ, Houck KA, Kavlock RJ, Knudsen TB, Martin MT, et al. 2013. Real-time growth kinetics measuring hormone mimicry for toxcast chemicals in t-47d human ductal carcinoma cells. Chem Res Toxicol 26:1097-1107.

Rusyn I, Sedykh A, Low Y, Guyton KZ, Tropsha A. 2012. Predictive modeling of chemical hazard by integrating numerical descriptors of chemical structures and short-term toxicity assay data. Toxicological Sciences 127:1-9.

Shah I, Wambaugh J. 2010. Virtual tissues in toxicology. Journal of Toxicology and Environmental Health Part B: Critical Reviews 13:314-328.

Sipes NS, Martin MT, Reif DM, Kleinstreuer NC, Judson RS, Singh AV, et al. 2011a. Predictive models of prenatal developmental toxicity from toxcast high-throughput screening data.

Toxicological Sciences 124:109-127.

Sipes NS, Padilla S, Knudsen TB. 2011b. Zebrafish: As an integrative model for twenty-first century toxicity testing. Birth Defects Research: Part C, Embryo Today 93:256-267.

Sipes NS, Martin MT, Kothiya P, Reif DM, Judson RS, Richard AM, et al. 2013. Profiling 976 toxcast chemicals across 331 enzymatic and receptor signaling assays. Chemical Research in Toxicology 26:878-895.

Skolness SY, Blanksma CA, Cavallin JE, Churchill JJ, Durhan EJ, Jensen KM, et al. 2013. Propiconazole inhibits steroidogenesis and reproduction in the fathead minnow (pimephales promelas). Toxicological Sciences 132(2):284-97.

Smith MT, Zhang L, McHale CM, Skibola CF, Rappaport SM. 2011. Benzene, the exposome and future investigations of leukemia etiology. Chemico-Biological Interactions 192:155-159.

Teschendorff AE, Widschwendter M. 2012. Differential variability improves the identification of cancer risk markers in DNA methylation studies profiling precursor cancer lesions. Bioinformatics 28:1487-1494.

Thayer KA, Heindel JJ, Bucher JR, Gallo MA. 2012. Role of environmental chemicals in diabetes and obesity: A national toxicology program workshop report. Environmental Health Perspectives 120:779-789.

Thienpont B, Tingaud-Sequeira A, Prats E, Barata C, Babin PJ, Raldua D. 2011. Zebrafish eleutheroembryos provide a suitable vertebrate model for screening chemicals that impair thyroid hormone synthesis. Environmental Science and Technology 45:7525-7532.

Thomas R, Phuong J, McHale CM, Zhang L. 2012. Using bioinformatic approaches to identify pathways targeted by human leukemogens. International Journal of Environmental Research and Public Health 9:2479-2503.

Thomas R, McHale CM, Lan Q, Hubbard AE, Zhang L, Vermeulen R, et al. 2013. Global gene expression response of a population exposed to benzene: A pilot study exploring the use of rna-sequencing technology. Environmental and Molecular Mutagenesis 54:566-573.

Thomas R, Hubbard A, McHale CM, Zhang L, Rappaport SM, Rothman N, et al. 2014. Characterization of changes in gene expression and biochemical pathways at low levels of benzene exposure. PloS One 9:e91828.

Thomas RS, Clewell HJ, Allen BC, Wesselkamper SC, Wang NC, Lambert JC, et al. 2011. Application of transcriptional benchmark dose values in quantitative cancer and noncancer risk assessment. Toxicological Sciences 120:194-205.

Thomas RS, Clewell HJ, 3rd, Allen BC, Yang L, Healy E, Andersen ME. 2012. Integrating pathway-based transcriptomic data into quantitative chemical risk assessment: A five chemical case study. Mutation Research 746:135-143.

Thomas RS, Himmelstein MW, Clewell HJ, 3rd, Yang Y, Healy E, Black MB, et al. 2013a. Cross-species transcriptomic analysis of mouse and rat lung exposed to chloroprene. Toxicological Sciences 131:629-640.

Thomas RS, Philbert MA, Auerbach SS, Wetmore BA, DeVito MJ, Cote I, et al. 2013b. Incorporating new technologies into toxicity testing and risk assessment: Moving from 21st

century vision to a data-driven framework. Toxicological Sciences.

Thomas RS, Wesselkamper SC, Wang NC, Zhao QJ, Petersen DD, Lambert JC, et al. 2013c. Temporal concordance between apical and transcriptional points of departure for chemical risk assessment. Toxicological Sciences 134:180-194.

Tice RR, Austin CP, Kavlock RJ, Bucher JR. 2013. Improving the human hazard characterization of chemicals: A Tox21 update. Environmental Health Perspectives 121:756-765.

Tollefsen KE, Scholz S, Cronin MT, Edwards SW, de Knecht J, Crofton K, et al. 2014. Applying adverse outcome pathways (AOPs) to support integrated approaches to testing and assessment (IATA). Regul Toxicol Pharmacol. pii: S0273-2300(14)00214-1.

Venkatapathy R, Wang NC. 2013. Developmental toxicity prediction. Methods in Molecular Biology 930:305-340.

Vermeulen R, Li G, Lan Q, Dosemeci M, Rappaport SM, Bohong X, et al. 2004. Detailed exposure assessment for a molecular epidemiology study of benzene in two shoe factories in china. The Annals of occupational hygiene 48:105-116.

Villeneuve D, Volz DC, Embry MR, Ankley GT, Belanger SE, Leonard M, et al. 2014. Investigating alternatives to the fish early-life stage test: A strategy for discovering and annotating adverse outcome pathways for early fish development. Environmental toxicology and chemistry / SETAC 33:158-169.

Vinken M. 2013. The adverse outcome pathway concept: A pragmatic tool in toxicology. Toxicology 312:158-165.

Wang N, Venkatapathy R, Bruce RM, Moudgal C. 2011. Development of quantitative structure-activity relationship (OSAR) models to predict the carcinogenic potency of chemicals. I. Using oral slope factor as a measure of carcinogenic potency. Regulatory Toxicology and Pharmacology 59:215-226.

Wang N, Jay Zhao Q, Wesselkamper SC, Lambert JC, Petersen D, Hess-Wilson JK. 2012a. Application of computational toxicological approaches in human health risk assessment. II. A tiered surrogate approach. Regulatory Toxicology and Pharmacology 63:10-19.

Wang N, Rice GE, Teuschler LK, Colman J, Yang RS. 2012b. An in silico approach for evaluating a fraction-based, risk assessment method for total petroleum hydrocarbon mixtures. Journal of Toxicology 2012:410143.

Warner CM, Gust KA, Stanley JK, Habib T, Wilbanks MS, Garcia-Reyero N, et al. 2012. A systems toxicology approach to elucidate the mechanisms involved in RDX species-specific sensitivity. Environmental Science and Technology 46:7790-7798.

Weiss JN, Karma A, MacLellan WR, Deng M, Rau CD, Rees CM, et al. 2012. "Good enough solutions" and the genetics of complex diseases. Circulation Research 111:493-504.

Wetmore BA, Wambaugh JF, Ferguson SS, Sochaski MA, Rotroff DM, Freeman K, et al. 2012. Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment. Toxicological Sciences 125:157-174.

Wetmore BA, Wambaugh JF, Ferguson SS, Li L, Clewell HJ, 3rd, Judson RS, et al. 2013. Relative impact of incorporating pharmacokinetics on predicting in vivo hazard and mode of

action from high-throughput in vitro toxicity assays. Toxicological Sciences 132:327-346.

Wright FA, Shabalin AA, Rusyn I. 2012. Computational tools for discovery and interpretation of expression quantitative trait loci. Pharmacogenomics 13:343-352.

Van Dyck E, Nazarov PV, Muller A, Nicot N, Bosseler M, Pierson S, KVan Moer K, Palissot V, Mascaux C, Knolle U, Ninane V, Nati R, Bremnes RM, Vallar L, Berchem G, Schlesser M. Bronchial airway gene expression in smokers with lung or head and neck cancer. Cancer Med. Apr 2014; 3(2): 322–336.

Wang X, Pittman GS, Bandele OJ, Bischof JJ, Liu G, Brothers JF 2nd, Spira A, Bell DA. Linking polymorphic p53 response elements with gene expression in airway epithelial cells of smokers and cancer risk. Hum Genet. 2014 Dec;133(12):1467-76.

Zeise L, Bois FY, Chiu WA, Hattis D, Rusyn I, Guyton KZ. 2013. Addressing human variability in next-generation human health risk assessments of environmental chemicals. Environmental Health Perspectives 121:23-31.

Table 1. Governmental Partners

- Army Corps of Engineers
- California Environmental Protection Agency, Office of Environmental Health Hazard Assessment
- European Chemicals Agency
- European Joint Research Commission
- Health Canada
- L'Institut National de l'Environnement Industriel et des Risques
- U.S. Food and Drug Administration and National Center for Toxicological Research
- U.S. Centers for Disease Control and Prevention, National Center for Environmental Health, and Agency for Toxic Substances and Disease Registry
- U.S. Department of Defense
- U.S. National Center for Advancing Translational Science
- U.S. National Institute of Environmental Health Sciences and National Toxicology Program
- U.S. National Institute for Occupational Safety and Health

Table 2. Prototype use of new scientific tools and techniques (adapted from Krewski et al. 2014)

Decision context category	Screening and Prioritization	Limited-scope assessments	Major-scope assessments
Hazard identification and dose-response assessment methods			
Quantitative structure activity relationship models	■	■	
Toxicity pathways analysis	■	■	■
High-throughput *in vitro* assays	■	■	■
High-content omics assays		■	■
Biomarkers of effect		■	■
Molecular and genetic population-based studies			■
Dosimetry and exposure assessment methods			
In vitro-to-*in vivo* extrapolation	■	■	
Pharmacokinetic models and dosimetry	■	■	■
Biomarkers of exposure		■	■
Cross-cutting assessment methods			
Adverse outcome pathways	■	■	■
Bioinformatics and computational biology	■	■	■
Systems biology	■	■	■
Functional genomics		■	■

Table 3. Possible characteristics of fit-for-purpose assessments matched to illustrative decision-context categories.

Characteristics	Illustrative decision-context categories		
	Screening and Prioritization	**Limited-scope assessments**	**Major-scope assessments**
Uses of NexGen assessments	Screening chemicals with no data other than QSAR or HT data, e.g. • Queuing for research, testing, or assessment • Urgent or emergency response	Generally nonregulatory decision-making, e.g. • Urban air toxics • Potential water contaminants • Hazardous waste and superfund chemicals • Urgent or emergency response	Often regulatory decision-making, e.g. • National risk assessments • Community risk assessment • Special problems of national concern
Data sources	EPA databases such as ACToR and ToxCast	NIH and EC databases (e.g. PubChem, BioSystems, Array Express, NHANES)	All sources of policy-relevant data
New data types (Each assessment type also uses the data types from the column to the left.)	QSAR, high-throughput *in vitro* screening assays, read across. AOP development	High-content assays, medium throughput assays, knowledge mined large data sets, AOP development	Molecular epidemiology, clinical and animal studies, AOP network development
Exposure paradigms of studies considered	*In vitro,* in *silico*	All relevant	*All relevant*
Metabolism in test systems	Little to none	Partial to intact	Intact
Incorporation of toxicokinetics	Reverse toxicokinetic models	Reverse toxicokinetics models, biomonitoring	Dosimetry and PK modeling, biomonitoring
Use of traditional *in vivo* data	*In vitro* assays anchored to pesticide registration and pharmaceutical data	None to limited, especially maybe used in AOP development	New data types augment traditional; traditional data remain basis for assessment at this time
Hazards	Nonspecific	Nonspecific to Identified	Identified
Potency metrics	Relative rankings based on QSAR or HT toxicity values	Relative rankings and toxicity values	Risk distributions, cumulative risks, community risks
Likely strength of evidence linking exposure to adverse effect	Suggestive	Suggestive to likely	Suggestive to known
Numbers of chemicals that can be assessed	10,000s	100s–1000s	100s
Time to conduct	Hours–Days	Hours–Weeks	Days–Years

assessment			

QSAR = quantitative structure activity relationship; HT = high throughput, EPA = U.S. Environmental Protection Agency, ACTor = Aggregated Computational Toxicology Resource (EPA), ToxCast = Toxicity Forecaster, NIH = National Institutes of Health, EC = European Community, NHANES = National Health and Nutrition Examination Survey, AOP = adverse outcome pathway

Figure 1. Three broad decision context categories are shown across the top (white type), below which are the seven "fit-for-purpose" prototypes developed for this effort (black type). PAHs = Polycyclic aromatic hydrocarbons; B[a]P = Benzo[a] pyrene.

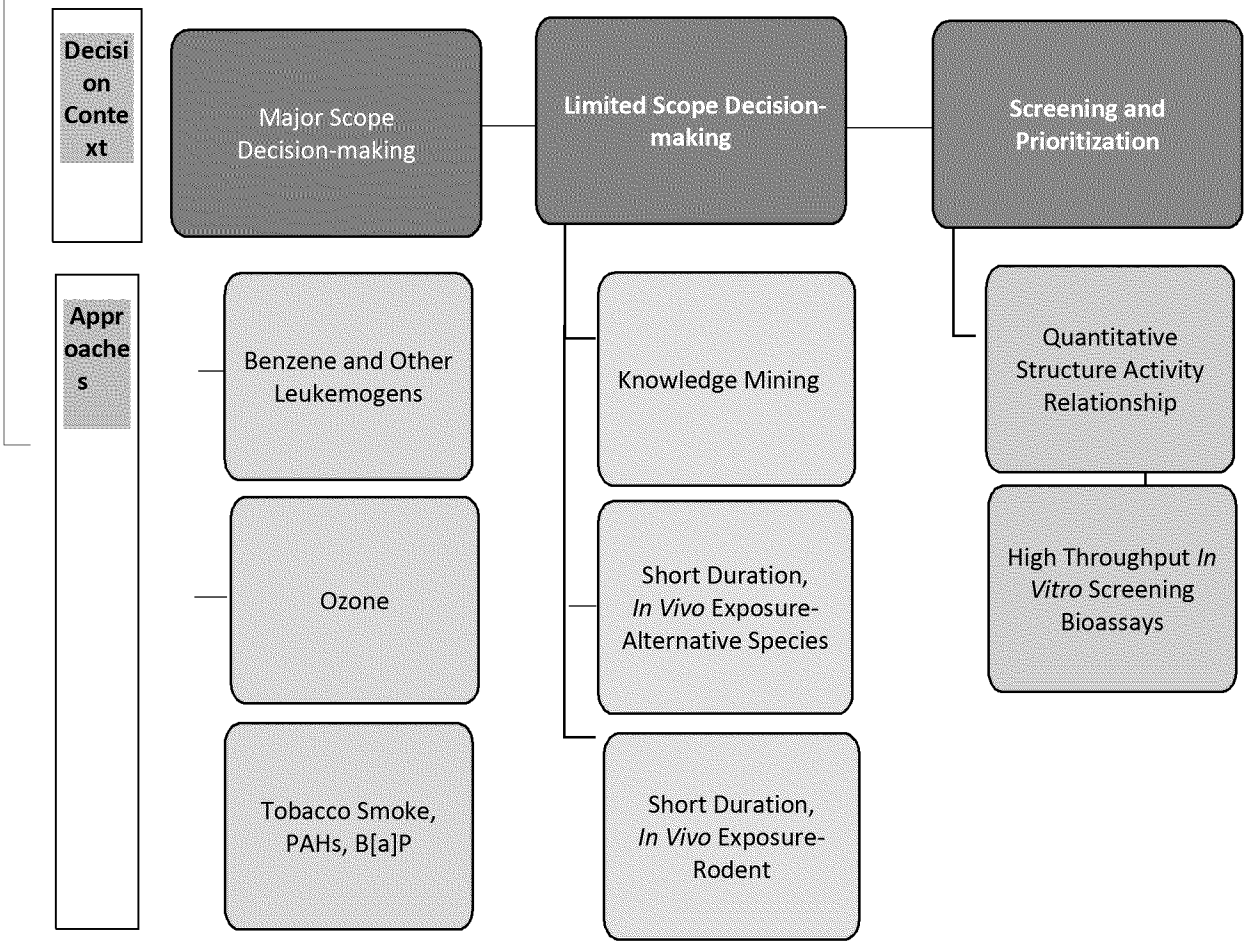
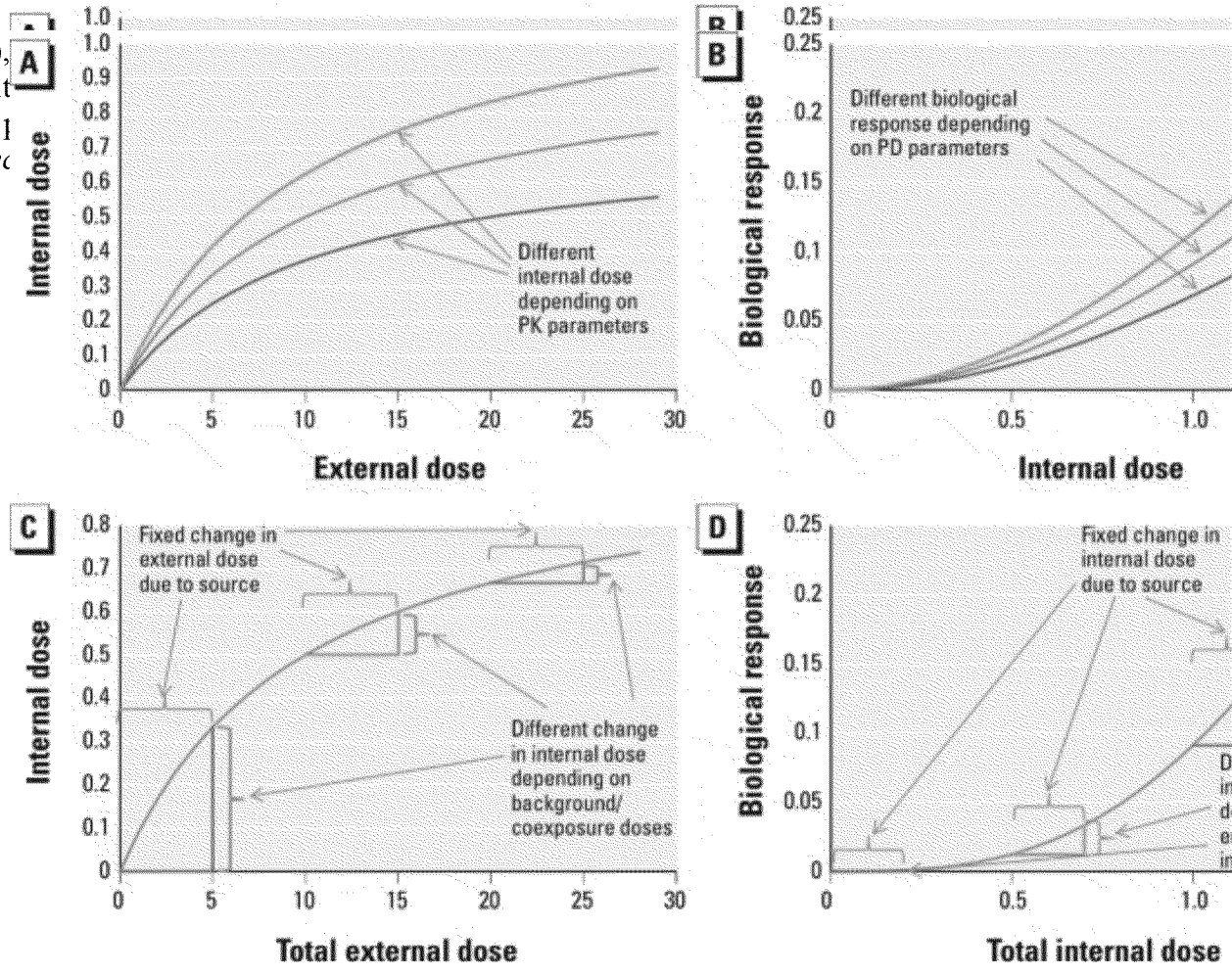
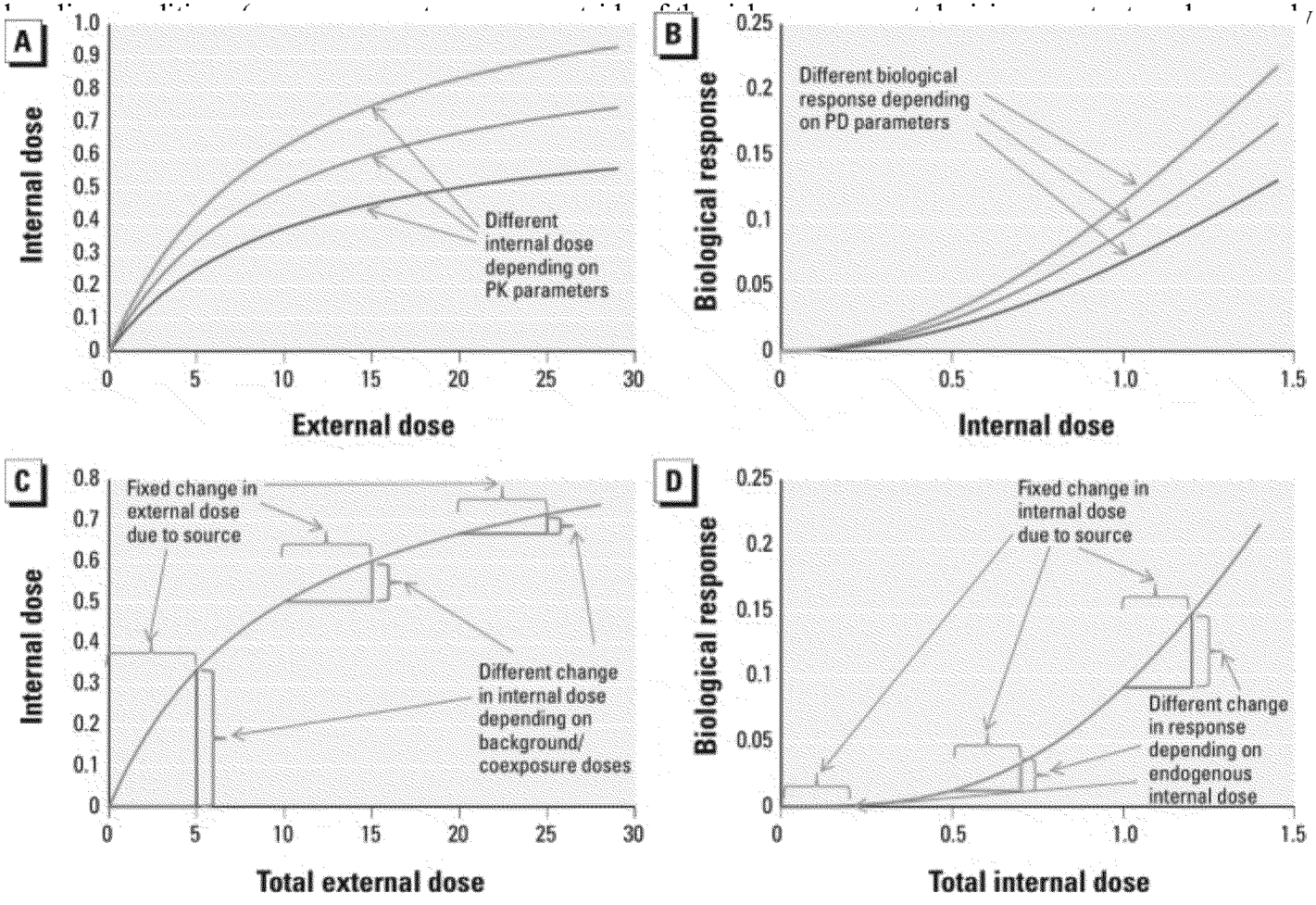


Figure 2. Effects of variability in (A) pharmacokinetics (PK), (B) pharmacodynamics (PD), (C) background/exposures, and (D) endogenous concentrations.



In (A) and (B),
baseline condit
endogenously
Health Perspec

Figure 2. Effects of variability in (A) pharmacokinetics (PK), (B) pharmacodynamics (PD), (C) background/exposures, and (D) endogenous concentrations.
In (A) and (B), individuals differ in PK or PD parameters. In (C) and (D), individuals have different initial



In (A) and (B), individuals differ in PK or PD parameters. In (C) and (D), individuals have different initial baseline conditions (e.g., exposure to sources outside of the risk management decisions context; endogenously produced compounds) (Zeise et al. 2013). Reproduced with permission from *Environmental Health Perspectives*.

Supplemental Table 1. List of Technical Papers in Association with the NexGen Report That Provide Additional Scientific Details

Preparation for Prototype Development	A framework for the next generation of risk science by Daniel Krewski, Margit Westphal, Mel Andersen, Greg Paoli, Weihsueh Chiu, Mustafa Al-Zoughool, Maxine Croteau, Lyle Burgoon, and Ila Cote (2014)
	Advancing the next generation of health risk assessment by Ila Cote, Paul Anastas, Linda Birnbaum, Becki Clark, David Dix, Stephen Edwards, and Peter Preuss (2012)
	Summary Report of Advancing the Next Generation of Risk Assessment Public Dialogue Conference by EPA (2011)
	Advancing the Next Generation (NexGen) of Risk Assessment: The Prototypes Workshop by EPA (2010)
Tier 3 Prototypes: Leukemia & Benzene, Lung Injury & Ozone, Liver Cancer & B[a]P/PAHs	Progress in assessing air pollutant risks from *in vitro* exposures: Matching ozone dose and effect in human airway cells by Gary Hatch, Kelly Duncan, David Diaz-Sanchez, Michael Schmitt, Andrew Ghio, Martha Carraway, John McKee, Lisa Dailey, Jon Berntsen, and Robert Devlin (2014)
	Ozone induces a pro-inflammatory response in primary human bronchial epithelial cells through MAP kinase activation without NF-κB activation by Sean McCullough, Kelly Duncan, Samantha Swanton, Lisa Dailey, David Diaz-Sanchez, and Robert Devlin (2014)
	Characterization of changes in gene expression and biochemical pathways at low levels of benzene exposure by Reuben Thomas, Alan Hubbard, Cliona McHale, Luoping Zhang, Stephen Rappaport, Qing Lan, Nathaniel Rothman, Kathryn Guyton, Roel Vermeulen, Jennifer Jinot, Babasaheb Sonawane, and Martyn Smith (2014)
	Temporal profile of gene expression alterations in primary human bronchial epithelial cells following *in vivo* exposure to ozone by Kelly Duncan, James Crooks, David Miller, Lyle Burgoon, Michael Schmitt, Stephen Edwards, David Diaz-Sanchez, and Robert Devlin (2013)
	IRIS Toxicological Review of Benzo[a]pyrene (Public Comment Draft). U.S. Environmental Protection Agency (2013), Washington, DC, EPA/635/R-13/138a-b
	Current understanding of the mechanism of benzene-induced leukemia in humans: Implications for risk assessment by Cliona McHale, Luoping Zhang, and Martyn Smith (2012)
	Benzene, the exposome and future investigations of leukemia etiology by Martyn Smith, Luoping Zhang, Cliona McHale, Christine Skibola, and Stephen Rappaport (2011)
	Global gene expression profiling of a population exposed to a range of benzene levels by Cliona McHale, Luoping Zhang, Qing Lan, Roel Vermeulen, Guilan Li, Alan Hubbard, Kristin Porter, Reuben Thomas, Christopher Portier, Min Shen, Stephen Rappaport, Songnian Yin, Martyn Smith, and Nathaniel Rothman (2010)
Tier 2 Prototypes Knowledge Mining Diabetes/ Obesity	Building associations between markers of environmental stressors and adverse human health impacts using frequent itemset mining by Shannon Bell and Stephen Edwards (2014)
	Systematic identification of interaction effects between genome- and environment-wide associations in type 2 diabetes mellitus by Chirag Patel, Rong Chen, Keiichi Kodama, John Ioannidis, and Atul Butte (2013)
	Data-driven integration of epidemiological and toxicological data to select candidate interacting genes and environmental factors in association with disease by Chirag Patel, Rong Chen, and Atul Butte (2012)
	Genetic variability in molecular responses to chemical exposure by Chirag Patel and Mark Cullen (2012)
	Role of environmental chemicals in diabetes and obesity: A National Toxicology Program workshop review by Kristina Thayer, Jerrold Heindel, John Bucher, and Michael Gallo (2012)
Tier 2 Prototypes Short-term in Vivo Non mammalian	Current perspectives on the use of alternative species in human health and ecological hazard assessments by Edward Perkins, Gerald Ankley, Kevin Crofton, Natàlia Garcia-Reyero, Carlie LaLone, Mark Johnson, Joseph Tietge, and Daniel Villeneuve (2013)

	Propiconazole inhibits steroidogenesis and reproduction in the fathead minnow (*Pimephales promelas*) by Sarah Skolness, Chad Blanksma, Jenna Cavallin, Jessica Churchill, Elizabeth Durhan, Kathleen Jensen, Rodney Johnson, Michael Kahl, Elizabeth Makynen, Daniel Villeneuve, and Gerald Ankley (2013)
	Zebrafish developmental screening of the ToxCast™ Phase I chemical library by Stephanie Padilla, Daniel Corum, Beth Padnos, Deborah Hunter, Andrew Beam, Keith Houck, Nisha Sipes, Nicole Kleinstreuer, Thomas Knudsen, David Dix, and David Reif (2012)
	A systems toxicology approach to elucidate the mechanisms involved in RDX species-specific sensitivity by Christopher Warner, Kurt Gust, Jacob Stanley, Tanwir Habib, Mitchell Wilbanks, Natàlia Garcia-Reyero, and Edward Perkins (2012)
Tier 2 Prototypes Short-term In Vivo Mammalian	Incorporating new technologies into toxicity testing and risk assessment: Moving from 21st century vision to a data-driven framework by Russell Thomas, Martin Philbert ,Scott Auerbach, Barbara Wetmore, Michael DeVito, Ila Cote, Craig Rowlands, Maurice Whelan, Sean Hays, Melvin Andersen, Bette Meek, Lawrence Reiter, Jason Lambert, Harvey Clewell III, Martin Stephens, Jay Zhao, Scott Wesselkamper, Lynn Flowers, Edward Carney, Timothy Pastoor, Dan Petersen, Carole Yauk, and Andy Nong (2013)
	Temporal concordance between apical and transcriptional points of departure for chemical risk assessment by Russell Thomas, Scott Wesselkamper, Nina Wang, Jay Zhao, Dan Peterson, Jason Lambert, Ila Cote, Longlong Yang, Eric Healy, Michael Black, Harvey Clewell III, Bruce Allen, and Melvin Andersen (2013)
	Integrating pathway-based transcriptomic data into quantitative chemical risk assessment: A five chemical case study by Russell Thomas, Harvey Clewell III, Bruce Allen, Longlong Yang, Eric Healy, and Melvin Andersen (2012)
	Application of transcriptional benchmark dose values in quantitative cancer and noncancer risk assessment by Russell Thomas, Harvey Clewell III, Bruce Allen, Scott Wesselkamper, Nina Wang, Jason Lambert, Janet Hess-Wilson, Jay Zhao, and Melvin Andersen (2011)
Tier 1 Prototypes Integration of QSAR and Various Biological Data Streams	Developmental toxicity prediction by Raghuraman Venkatapathy and Nina Wang (2013)
	Predictive QSAR modeling: Methods and applications in drug discovery and chemical risk assessment by Alexander Golbraikh, Xiang Simon Wang, Hao Zhu, and Alexander Tropsha (2012)
	Predictive modeling of chemical hazard by integrating numerical descriptors of chemical structures and short-term toxicity assay data by Ivan Rusyn, Alexander Sedykh, Yen Low, Kathryn Guyton, and Alexander Tropsha (2012)
	Application of computational toxicological approaches in human health risk assessment I. A tiered surrogate approach by Nina Wang, Jay Zhao, Scott Wesselkamper, Jason Lambert, Dan Petersen, and Janet Hess-Wilson (2012)
	An *in silico* approach for evaluating a fraction-based, risk assessment method for total petroleum hydrocarbon mixtures by Nina Wang, Glenn Rice, Linda Teuschler, Joan Colman, and Raymond Yang (2012)
	Development of quantitative structure-activity relationship (QSAR) models to predict the carcinogenic potency of chemicals. II. Using oral slope factor as a measure of carcinogenic potency by Nina Wang, Raghuraman Venkatapathy, Robert Mark Bruce, and Chandrika Moudgal (2011)
Tier 1 Prototypes High-throughput Screening	*In vitro* and modelling approaches to risk assessment from the U.S. Environmental Protection Agency ToxCast programme by Richard Judson, Keith Houck, Matt Martin, Thomas Knudsen, Russell Thomas, Nisha Sipes, Imran Shah, John Wambaugh, and Kevin Crofton (2014)
	Perspectives on validation of high-throughput assays supporting 21st century toxicity testing by Richard Judson, Robert Kavlock, Matthew Martin, David Reif, Keith Houck, Thomas Knudsen, Ann Richard, Raymond Tice, Maurice Whelan, Menghang Xia, Ruili Huang, Christopher Austin, George Daston, Thomas Hartung, John Fowle III, William Wooge, Weida Tong, and David Dix (2013)
	Estimating toxicity-related biological pathway altering doses for high-throughput chemical risk assessment by Richard Judson, Robert Kavlock, Woodrow Setzer, Elaine Cohen Hubal, Matthew Martin, Thomas Knudsen, Keith Houck, Russell Thomas, Barbara Wetmore, and David Dix (2011)

Key Risk Assessment Issues	The role of advanced biological methods and data in regulatory rationality of risk-based regulatory decisions by Douglas Crawford-Brown (2013)
	Incorporating new technologies into toxicity testing and risk assessment: Moving from 21st century vision to a data-driven framework by Russell Thomas, Martin Philbert, Scott Auerbach, Barbara Wetmore, Michael DeVito, Ila Cote, Craig Rowlands, Maurice Whelan, Sean Hays, Melvin Andersen, Bette Meek, Jason Lambert, Harvey Clewell III, Martin Stephens, Jay Zhao, Scott Wesselkamper, Lynn Flowers, Edward Carney, Timothy Pastoor, Dan Petersen, Carol Yauk, and Andy Nong (2013)
	Addressing human variability in next-generation human health assessments of environmental chemicals by Lauren Zeise, Frederic Bois, Weihsueh Chiu, Dale Hattis, Ivan Rusyn, and Kathryn Guyton (2012)
	Quantitative high-throughput screening for chemical toxicity in a population-based *in vitro* model by Eric Lock, Nour Abdo, Ruili Huang, Menghang Xia, Oksana Kosyk, Shannon O'Shea, Yi-Hui Zhou, Alexander Sedykh, Alexander Tropsha, Christopher Austin, Raymond Tice, Fred Wright, and Ivan Rusyn (2012)
	Predicting later-life outcomes of early-life exposures by Kim Boekelheide, Bruce Blumberg, Robert Chapin, Ila Cote, Joseph Graziano, Amanda Janesick, Robert Lane, Karen Lillycrop, Leslie Myatt, Christopher States, Kristina Thayer, Michael Waalkes, and John Rogers (2012)
	In vitro screening for population variability in chemical toxicity by Shannon O'Shea, John Schwarz, Oksana Kosyk, Pamela Ross, Min Jin Ha, Fred Wright, and Ivan Rusyn (2011)

Supplemental Table 2: An Illustrative Framework for Evidence Integration Focusing on New Data Types Presented in the NexGen Report. This causal determination framework illustrates how evidence integration and inferences about causality could be made using new data types. The left column summarizes the prototype results, the middle column presents evidence for causality exemplified by the prototypes, and the right column illustrates how such prototypic evidence might be integrated and weighed. The first set of prototypes is unique in that the prototypes have known human health effects and well-documented public health risks. For these prototypes, the "Evidence Integration" column evaluates how successful new data types were in predicting known outcomes. Criteria for study selection, evidence integration and causal determination considered here are discussed in McConnell et al, 2014, Meek et al 2014, NRC 2014, USDHHA 2014, EPA 2013a; Rhomberg et al. 2013. Modifications of the Bradford-Hill criteria (e.g. consistency, coherence, biologic plausibility) continued to be useful in the evaluation of new data types. As presented in this Table, confidence in causality ranges from suggestive to likely, largely based on the understanding of the biologic context in which new data types are embedded. "Likely" is generally for the limited number of cases where the new data types are well anchored to adverse outcomes by a combination of observational and experimental data, and include mechanistic and systems biology understanding. Biologic context need not be chemical specific but can be derived from disease/disorder specific knowledge or from analogy with related chemicals. In practice, for the near-term much of the new datatypes are anticipated to be suggestive and most appropriated for screening and prioritization and, perhaps, limited-scope assessments. Of note is that, contrary to traditional approaches, some new approaches can be used to estimate relative potencies or toxicity values in the absence of clearly identified hazards. Major assessments are anticipated to be augmented by new data types but, for the near term, continue to be based on traditional data. To simplify the table, similar prototypes with shared attributes are aggregated where possible.

Supplemental Table 2: An Illustrative Framework for Evidence Integration Focusing on New Data Types Presented in the NexGen Report.		
Prototypes	**Evidence for Causality**	**Evidence Integration**

Tier 3	**Molecular epidemiology and clinical** studies: • Illustrated that new data types (when properly collected, analyzed, and reported) appear to provide results comparable to robust, traditional human data, and could be used, when linked to mechanistic information, to: (1) evaluate potential hazard posed by chemicals with no or limited traditional data, (2) augment traditional assessments, or (3) better inform traditional risk assessment issues, such as human variability and susceptibility, cross-species and in vivo to in vitro comparisons, cumulative risk, and low exposure-dose-response relationships.	• **Specific pattern alterations in molecular events appear consistently and strongly associated with known intermediate events and known hazards at environmental exposure levels** (EPA 2014b, Thomas R et al. 2014, McCullough et al. 2014, EPA 2013a, McHale et al. 2009, 2012, Smith et al. 2011). Data for tobacco smoke was reported in the Gene Expression Omnibus (GEO) or ArrayExpress (AE). Also see Cooper et al. 2014, Van Dyke et al. 2014, and Wang et al. 2014). • Exposure-dose was measured for benzene and ozone using urinary biomarkers and radiolabeled ozone, respectively (Hatch et al. 2014, Thomas R et al. 2014, and McHale et al. 2009). Tobacco smoking exposures were self- reported substantially increasing uncertainty for exposure-dose-response characterization and highlighting the need for accurate exposure characterization (EPA 2014b). • Dose-dependent alterations are observed in concomitantly collected molecular events and adverse effects, in the range of environmental exposure (benzene and ozone). Some molecular pathways are altered at all concentrations; other molecular and toxicological effects emerge with increasing dose. Molecular patterns which occur consistently across all concentrations appear preferable as biomarkers (Hatch et al. 2014, Thomas et al. 2014, McCullough et al. 2014, and EPA 2013a). • Pharmacological interventions has been shown to modify identified AOPs, and, concomitantly, the incidence or severity of the adverse outcomes (McCullough et al. 2014, Hatzimichael and Crook 2013, Cooper et al. 2014).	**Suggestive to likely:** Evidence is consistent, coherent, and biologically plausible that the observed molecular events are causally related to adverse effects Implications based on comparisons to robust traditional risk assessments: • For benzene and ozone, identified molecular events are likely causally related to known adverse outcomes in a dose-dependent fashion. Mechanistic links between molecular events, intermediate effects and adverse outcomes are well understood. Pharmacologic intervention that blocks implicated pathways also blocks or ameliorates adverse effects. • In comparison, the molecular data for PAH are considered suggestive of a causal association between PAH and lung cancer due to a lack of an observed exposure-dose-response relationship (likely due to uncertainties in exposure characterization), and data quality, analysis, and reporting limitations. Only ~8% studies in GEO and AE met study selection criteria.

		• Additional evidence for the involvement of specific pathways in disease is provided by identification of naturally occurring human gene variants in the AOP network that alter susceptibility and risks (Cooper et al. 2014, Vawda et al. 2014, Wang et al. 2014, Moreno-Macias et al. 2013, Hatzimichael and Crook 2013, Zhuo et al. 2012, Sille et al. 2012, North et al. 2011, Shen et al. 2011, Smith et al. 2011, Schlenk et al. 2008). • Adverse outcome pathway (AOP) networks are also disrupted by other chemical and nonchemical stressors known to alter the incidence of the specific disease/disorder under consideration; thus, AOP networks provide a tool for evaluating cumulative risks based on mechanistic commonalities (IARC 2012, Thomas et al. 2012, Smith et al. 2011) • Supporting data is provided by multiple molecular epidemiology and clinical studies and chronic animal bioassays and coherence with other systems biology data (NIH BioSystems: acute myeloid leukemia, lung cancer (small cell, non-small cell); Comparative Toxicogenomics Database: PAHs and cancer: EPA 2013a,b: BaP and cancer, ozone and respiratory disease). • While species and *in vitro* differences exist, these examples provide consistent, coherent biologically plausible data linking specific omic alterations with specific diseases.	**Suggestive vs. likely:** In general, molecular data alone associated with adverse outcomes are expected to be only suggestive or inadequate for causal determination. To rise to likely, the following are currently, generally necessary: multiple, consistent, high-quality observational studies with similar results; understanding of the cascade of events between molecular events to adverse outcomes, and experimental evidence showing that reversal of pathway alterations blocks or ameliorates adverse outcome; or naturally occurring experiments where gene variants alter incidence or characteristic of disease. Important variables such as experimental paradigm (e.g., *in vivo* vs. *in vitro*), cell type, tissue type, and species also require consideration. Suggestive data are likely to be most useful for hypothesis generation, discovery, screening and prioritization, and potential augmentation of traditional data.

Tier 2	**Knowledge mining and meta-analysis** prototype illustrated how large searchable databases can be used to identify, organize, integrate, and analyze existing data in an automated (computerized) fashion to discover new insights into public health risks.	• **Knowledge mining and meta-analysis discovered associations between known exposures (biomonitoring) to several environmental agents** e.g. metals and persistent organics chemical) with prediabetes/diabetes using the Centers for Disease Control and Prevention's National Health Assessment Examination Survey. Human tissue biomarker and clinical outcome data are from the same individuals (EPA 2014b, Patel 2012a, 2013). • Supporting data are found elsewhere in the literature (Comparative Toxicogenomic Database chemical and Type II Diabetes associations 2014, Thayer et al. 2012). • No systems biology context or AOP is available data from NHANES. Information on biologic context is available from the literature but currently not easily accessible in high or medium throughput approaches (Audouze et al. 2013, Inadera 2013). • Prototype also explored possible links among site specific chemical exposures, ethnicity, genetic variants and diabetes risks (Dimas et al. 2014, EPA 2014b, Patel and Cullen 2012).	**Suggestive**: Could rise to likely with the types of supporting data noted above under "Suggestive vs. likely." Associative data generally most useful for hypothesis generation.
	Short-duration *in vivo* exposure bioassays data use in either alternative or rodents species is illustrated in two prototypes.	**Short-duration, in vivo exposure bioassays - alternative (nonmammalian) species prototype included example Zebrafish developmental assay results, characterization of thyroid specific mechanisms, and predictive modeling of complex dose-response phenomena.** Padilla et al. 2012, reported AC50s in a Zebrafish developmental assay for 305 chemicals. Potencies for individual chemicals and chemical classes were shown to range over several orders of magnitude (1 nM–80 µM). For certain classes of chemicals, 80-100% of the chemicals in a class tested positive (embryo death or structural defects). Perkins et al. 2013 illustrated the use of alternative species to help articulate mechanisms showing an AOP network for thyroid disruption with example toxicants and alternative models applicable to both human and ecological hazard assessment. Also, discussed are how predictive models coupled to mechanistic understanding can be used to better characterize dose-response (Sipe et al 2011b), circadian variations (Eisenberg et al. 2008) and exposure window-response relationships (Dewoskin et al. 2014). For more details on alternative species bioassays see Villeneuve et al. 2014, Perkins et al. 2013, Ankely and Gray 2012).	**Suggestive to likely** for consistent, coherent, biologically plausible adverse phenotypic outcome data from nonmammalian, vertebrate species. Confidence is generally higher for evolutionarily conserved processes. **Suggestive:** For transcriptomics changes correlated to adverse outcomes studies and coupled to AOPs. High-content assays with measurable adverse outcomes (e.g., zebrafish developmental assay) generally have greater evidentiary weight than initiating event assays (e.g. transcriptomic assays). Some systems biology context is needed for limited scope assessments for human risk, e.g. cross species conservation, AOPs. Alternative species outcome data alone are sufficient for ecologic risk assessment. Cross-species extrapolation and subchronic measurement of indicators introduces additional uncertainties as compared to human data discussed above data. Reverse toxicokinetic models are needed to estimate equivalent human doses.

		Short-duration, in vivo exposure bioassays - rodents prototype correlated transcriptomic alterations with adverse outcomes, as determined in traditional bioassays for 10 chemicals (Thomas RS et al. 2011, 2012, 2013, 2014). Cconsistency of the correlation between transcriptional changes and adverse effects across different exposure periods was also demonstrated (5 days to 13 weeks) (Thomas, R. S. et al. 2013d). Ttranscriptional changes appeared at somewhat lower concentrations than traditional effects. Transcriptomic studies alone cannot predict specific hazards but may be useful to relatively ranking chemical potencies to induce biologic alterations that may proceed adverse outcomes. In general transcriptomic data needs some biologic context (e.g. AOPs) to increase confidence of biologic significance.	
Tier 1	**QSAR and molecular docking models** are used to generate potency estimates and, with less confidence, hazards. Read-across is also considered (i.e. filling data gaps for data poor chemicals by analogy with structurally related more data rich chemicals).	**QSAR models can predict chemical-specific toxicity values based on chemical inherent properties for a number of data poor chemicals.** • Models are developed based on chemical structures and known outcomes for data rich chemicals. • OECD is harmonizing international use of QSAR hazard models and read-across in the OECD QSAR toolbox (OECD 2014 b,c,e) • Often the consensus of a suite of appropriate models is the preferred approach. • Often models better predict potency than specific effects. • Issues exist around characterizing the uncertainty in QSAR and related read-across approaches, and in the transparency of some models (see Ball et al. 2014; Patlewicz et al. 2013a)].	**Suggestive to Likely:** TopKat Model predictions of potency when model is appropriate for chemicals evaluated; not generally predictive of dose-response for specific hazards; does generate a LOAEL for a subset of the data poor chemicals that meet confidence criteria. Additional OECD models and read-across can improve confidence in hazard characterization.
	High-throughput, *in vitro* bioassays and virtual tissue models are discussed.	**High-throughput *in vitro* assays** based on biological process disruptions are interpreted in a systems biology and AOP context, and associated with adverse outcomes (EPA 2014b, Judson et al. 2014, Attene-Ramos et al. 2013, Tice et al. 2013). **Virtual tissue modeling** provides additional tools for evaluating data limited chemicals (Knudsen et al. 2013).	**Suggestive:** When coupled with understanding of the AOP(s). Could rise to likely with the types of supporting data noted above under "Suggestive vs. likely".